Amendments to the Specification:

On page 4, please amend the paragraph starting on line 13 as follows:
-- Fig. 2 shows Scatterplots of the four different criteria used by SMAR ~~Scan®~~ SCAN and the AT-content with human MARs from SMARt DB. –

-

On page 4, please amend the paragraph starting on line 21 as follows:
-- Fig. 4 shows SMAR ~~Scan®~~ SCAN predictions on human chromosome 22 and on shuffled chromosome 22. Top plot : Average number of hits obtained by SMAR ~~Scan®~~ SCAN with five: rubbled, scrambled, shuffled within nonoverlapping windows of 10 bp, order 1 Markov chains model and with the native chromosome 22. Bottom plot: Average number of MARs predicted by SMAR ~~Scan®~~ SCAN in five: rubbled, scrambled, shuffled within nonoverlapping windows of 10 bp, order 1 Markov chains model and with the native chromosome 22. --

On page 5, please amend the paragraph starting on line 24 as follows:
-- Fig. 10 shows the distribution of putative transcription factor binding sites within the 5'-cLysMAR. Large arrows indicate the position of the CUE elements as identified with SMAR ~~Scan®~~ SCAN.--

On page 5, please amend the paragraph starting on line 51 as follows:
-- Fig. 15 shows comparative performance of SMAR prediction algorithms exemplified by region WP18A10A7. (A) SMAR ~~Scan®~~ SCAN analysis was performed with default settings. (B) SIDD analysis (top curve and left-hand side scale), and the attachment of several DNA fragments to the nuclear matrix in vitro (bar-graph, right-hand side scale) was taken from Goetze et al (Goetze S, Gluch A, Benham C, Bode J, "Computational and in vitro analysis of destabilized DNA regions in the interferon gene cluster: potential of predicting functional gene domains." *Biochemistry*, 42:154-166, 2003).--

On page 6, please amend the paragraph starting on line 12 as follows:
-- Fig. 17 represents the scatterplot for the 1757 S/MAR sequences of the AT (top) and TA (bottom) dinucleotide percentages versus the predicted DNA bending as computed by SMAR ~~Scan®~~ SCAN.--

On page 7, please amend the paragraph starting on line 24 as follows:
-- The sequences SEQ ID Nos 1 to 23 have been identified by scanning human chromosome 1 and 2 using SMAR ~~Scan®~~ SCAN, showing that the identification of novel MAR sequences is feasible using the tools reported thereafter whereas SEQ ID No 24 to 27 have been identified by scanning the complete human genome using the combined SMAR ~~Scan®~~ SCAN method. --

On page 12, please amend the paragraph starting on line 1 as follows:
-- The bioinformatic tool used for the present method is preferably, SMAR ~~Scan®~~ SCAN, which contains algorithms developed by Gene Express and based on Levitsky *et al.*, 1999. These algorithms recognise profiles, based on dinucleotides weight-matrices, to compute the theoretical values for conformational and physicochemical properties of DNA. --

On page 12, please amend the paragraph starting on line 7 as follows:
-- Preferably, SMAR ~~Scan®~~ SCAN uses the four theoretical criteria also designated as DNA sequence features corresponding to DNA bending, major groove depth and minor groove width potentials, melting temperature in all possible combination, using scanning windows of variable size (see Fig. 3). For each function used, a cut-off value has to be set. The program returns a hit every time the computed score of a given region is above the set cut-off value for all of the chosen criteria. Two data output modes are available to handle the hits, the first (called "profile-like") simply returns all hit positions on the query sequence and their corresponding values for the different criteria chosen. The second mode (called "contiguous hits") returns only the positions of several contiguous hits and their corresponding sequence. For this mode, the minimum number of contiguous hits is another cut-off value that can be set, again with a tunable window size. This second mode is the default mode of SMAR ~~Scan®~~ SCAN. Indeed, from a semantic point of view, a hit is considered as a core-unwinding element (CUE), and a cluster of CUEs accompanied by clusters of binding sites for relevant proteins is considered as a MAR. Thus, SMAR ~~Scan®~~ SCAN considers only several contiguous hits as a potential MAR. --

On page 12, please amend the paragraph starting on line 23 as follows:
-- To tune the default cut-off values for the four theoretical structural criteria, experimentally validated MARs from SMARt DB were used. All the human MAR sequences from the database were retrieved and analyzed with SMAR ~~Scan®~~ SCAN using the "profile-like" mode with the four criteria and with no set cut-off value. This allowed the setting of each function for every position

of the sequences. The distribution for each criterion was then computed according to these data (see Fig. 1 and 3). --

On page 12, please amend the paragraph starting on line 31 as follows:
-- The default cut-off values of SMAR ~~Scan®~~ SCAN for the bend, the major groove depth and the minor groove width were set at the average of the 75th quantile and the median. For the melting temperature, the default cut-off value should be set at the 75th quantile. The minimum length for the "contiguous-hits" mode should be set to 300 because it is assumed to be the minimum length of a MAR (see Fig. 8 and 9). However, one skilled in the art would be able to determine the cut-off values for the above-mentioned criteria for a given organism with minimal experimentation. --

On page 13, please amend the paragraph starting on line 13 as follows:
-- In case SMAR ~~Scan®~~ SCAN is envisaged to perform, for example, large scale analysis, then, preferably, the above-mentioned method further comprises at least one filter predicting DNA binding sites for DNA transcription factors in order to reduce the computation. --

On page 13, please amend the paragraph starting on line 17 as follows:
-- The principle of this method combines SMAR ~~Scan®~~ SCAN to compute the structural features as described above and a filter, such as for example, the pfsearch, (from the pftools package as described in Bucher P, Karplus K, Moeri N, and Hofmann K, "A flexible search technique based on generalized profiles", *Computers and Chemistry* , 20:324, 1996) to predict the binding of some transcription factors. --

On page 13, please amend the paragraph starting on line 26 as follows:
-- This combined method uses the structural features of SMAR ~~Scan®~~ SCAN and the predicted binding of specific transcription factors of the filter that can be applied sequentially in any order to select MARs, therefore, depending on the filter is applied at the beginning or at the end of the method. --

On page 13, please amend the paragraph starting on line 42 as follows:
-- The combined method is actually a "wrapper" written in Perl for SMAR ~~Scan®~~ SCAN and, in case the pfsearch is used as a filter, from the pftools. The combined method performs a twolevel

processing using at each level one of these tools (SMAR ~~Scan®~~ SCAN or filter) as a potential "filter", each filter being optional and possible to be used to compute the predicted features without doing any filtering. --

On page 13, please amend the paragraph starting on line 48 as follows:
-- If SMAR ~~Scan®~~ SCAN is used in the first level to filter subsequences, it has to be used with the "all the contiguous hits" mode in order to return sequences. If the pfsearch is used in the first level as first filter, it has to be used with only one profile and a distance in nucleotide needs to be provided. This distance is used to group together pfsearch hits that are located at a distance inferior to the distance provided in order to return sequences; The combined method launches pfsearch, parses its output and returns sequences corresponding to pfsearch hits that are grouped together according to the distance provided. Then whatever the tool used in the first level, the length of the subsequences thus selected can be systematically extended at both ends according to a parameter called "hits extension". --

On page 14, please amend the paragraph starting on line 6 as follows:
-- The second and optional level can be used to filter out sequences (already filtered sequences or unfiltered input sequences) or to get the results of SMAR ~~Scan®~~ SCAN and/or pfsearch without doing any filtering on these sequences. If the second level of combined method is used to filter, for each criteria considered cutoff values (hit per nucleotide)need to be provided to filter out those sequences (see Fig. 20). –

On page 14, please amend the paragraph starting on line 12 as follows:
-- Another concern of the present invention is also to provide a method for identifying a MAR sequence comprising at least one filter detecting clusters of DNA binding sites using profiles or weightmatrices. Preferably, this method comprises two levels of filters and in this case, SMAR ~~Scan®~~ SCAN is totally absent from said method. Usually, the two levels consist in pfsearch. --

On page 23, please amend line 4 as follows:

-- ***Example 1: SMAR ~~Scan®~~ SCAN and MAR sequences*** --

On page 23, please amend the paragraph starting on line 6 as follows:

-- A first rough evaluation of SMAR ~~Scan®~~ SCAN was done by analyzing experimentally defined human MARs and non-MAR sequences. As MAR sequences, the previous results from the analysis of human MARs from SMARt Db were used to plot a density histogram for each criterion as shown in Fig. 1. Similarly, non-MAR sequences were also analyzed and plotted. As non-MAR sequences, all Ref-Seq-contigs from the chromosome 22 were used, considering that this latter was big enough to contain a negligible part of MAR sequences regarding the part of non-MAR sequences. --

On page 23, please amend the paragraph starting on line 47 as follows:
-- All RefSeq contigs from the chromosome 22 were analyzed by SMAR ~~Scan®~~ SCAN using the default settings this time. The result is that SMAR ~~Scan®~~ SCAN predicted a total of 803 MARs, their average length being 446 bp, which means an average of one MAR predicted per 42 777 bp. The total length of the predicted MARs corresponds to 1% of the chromosome 22 length. The AT-content of the predicted regions ranged from 65,1% to 93.3%; the average AT-content of all these regions being 73.5%. Thus, predicted MARs were AT-rich, whereas chromosome 22 is not AT-rich (52.1% AT). --

On page 24, please amend the paragraph starting on line 4 as follows:
-- SMARTest was also used to analyze the whole chromosome 22 and obtained 1387 MAR candidates, their average length being 494 bp representing an average of one MAR predicted per 24 765 bp. The total length of the predicted MARs corresponds to 2% of the chromosome 22. Between all MARs predicted by the two softwares, 154 predicted MARs are found by both programs, which represents respectively 19% and 11% of SMAR ~~Scan®~~ SCAN and SMARTest predicted MARs. Given predicted MARs mean length for SMAR ~~Scan®~~ SCAN and SMARTest, the probability to have by chance an overlapping between SMAR ~~Scan®~~ SCAN and SMARTest predictions is 0.0027% per prediction. --

On page 24, please amend the paragraph starting on line 14 as follows:
-- To evaluate the specificity of SMAR ~~Scan®~~ SCAN predictions, SMAR ~~Scan®~~ SCAN analyses were performed on randomly shuffled sequences of the chromosome 22 (Fig. 4). Shuffled sequences were generated using 4 different methods: by a segmentation of the chromosome 22 into nonoverlapping windows of 10 bp and by separately shuffling the nucleotides in each window; by "scrambling" which means a permutation of all nucleotides of the chromosome; by

"rubbling" which means a segmentation of the chromosome in fragments of 10 bp and a random assembling of these fragments and finally by order 1 Markov chains, the different states being the all the different DNA dinucleotides and the transition probabilities between these states being based on the chromosome 22 scan. For each shuffling method, five shuffled chromosome 22 were generated and analyzed by SMAR ~~Scan®~~ SCAN using the default settings. Concerning the number hits, an average of 3 519 170 hits (sd: 18 353) was found for the permutated chromosome 22 within nonoverlapping windows of 10 bp, 171 936,4 hits (sd: 2 859,04) for the scrambled sequences and 24 708,2 hits (sd: 1 191,59) for the rubbled chromosome 22 and 2 282 hits in average (sd: 334,7) for the chromosomes generated according to order 1 Markov chains models of the chromosome 22, which respectively represents 185% (sd: 0.5% of the mean), 9% (sd: 1.5%), 1% (sd: 5%) and 0.1% (sd: 15%) of the number of hits found with the native chromosome 22. For the number of MARs predicted, which thus means contiguous hits of length greater than 300, 1 997 MARs were predicted with the shuffled chromosome 22 within windows of 10 bp (sd: 31.2), only 2.4 MARs candidates were found in scrambled sequences (sd: 0.96) and none for the rubbled and for the sequences generated according to Markov chains model, which respectively represents 249% and less than 0.3% of the number of predicted MARs found with the native chromosome 22. These data provide indications that SMAR ~~Scan®~~ SCAN detects specific DNA elements which organization is lost when the DNA sequences are shuffled. --

On page 24, please amend line 6 and 7 as follows:

-- **Example 4: Analysis of known matrix attachment regions in the Interferon locus with SMAR ~~Scan®~~ SCAN** --

On page 24, please amend the paragraph starting on line 45 as follows:

-- The relevance of MAR prediction by SMAR ~~Scan®~~ SCAN was investigated by analyzing the recently published MAR regions of the human interferon gene cluster on the short arm of chromosome 9 (9p22). Goetze et al. (already cited) reported an exhaustive analysis of the WP18A10A7 locus to analyze the suspected correlation between BURs (termed in this case stress-induced duplex destabilization or SIDD) and *in vitro* binding to the nuclear matrix (Fig. 9, lower part). Three of the SIDD peaks were in agreement with the *in vitro* binding assay, while others did not match matrix attachment sites. Inspection of the interferon locus with SMAR ~~Scan®~~ SCAN (Fig. 9, top part) indicated that three majors peaks accompanied by clusters of

SATB1, NMP4 and MEF2 regulators binding sites correlated well with the active MARs. Therefore, we conclude that the occurrence of predicted CUEs and binding sites for these transcription factors is not restricted to the *cLys*MAR but may be a general property of all MARs. These results also imply that the SMAR ~~Scan®~~ SCAN program efficiently detects MAR elements from genomic sequences. --

On page 25, please amend the paragraph starting on line 6 as follows:
-- **Example 5: Accuracy of SMAR ~~Scan®~~ SCAN prediction and comparison with other predictive tools** --

On page 27, please amend line 3 as follows:
-- Table 1: Evaluation of SMAR ~~Scan®~~ SCAN accuracy --

On page 27, please amend the paragraph starting on line 5 as follows:
-- Six different genomic sequences, three plant and three human sequences, for which experimentally defined MARs are known, were analyzed with MAR-Finder, SMARTest and SMAR ~~Scan®~~ SCAN. True positive matches are printed in bold, minus (-) indicates false negative matches. Some of the longer experimentally defined MARs contained more than one in silico prediction, each of them was counted as true positive match. Therefore, the number of true in silico predictions is higher than the number of experimentally defined MARs found. Specificity is defined as the ratio of true positive predictions, whereas sensitivity is defined as the ratio of experimentally defined MARs found. * AT-rich rule excluded using MAR-Finder. --

On page 27, please amend the paragraph starting on line 15 as follows:
-- SMARTest predicted 28 regions as MARs, 19 (true positives) of these correlate with experimentally defined MARs (specificity: 68%) whereas 9 (32%) are located in non-MARs (false positives). As some of the longest experimentally determined
MARs contains more than one in silico prediction, the 19 true positives correspond actually to 14 different experimentally defined MARs (sensitivity: 38%). MARFinder predicted 25 regions as MARs, 20 (specificity: 80%) of these correlate with experimentally defined MARs corresponding to 12 different experimentally defined MARs (sensitivity: 32%). SMAR ~~Scan®~~ SCAN predicted 22 regions, 17 being true positives (specificity: 77%) matching 14 different experimentally defined MARs (sensitivity: 38%). --

On page 27, please amend line 29 and 30 as follows:

-- **Example 6: Analyses of the whole genome using the combined method (SMAR ~~Scan®~~ SCAN -pfsearch)** --

On page 27, please amend the paragraph starting on line 32 as follows:

-- In order to test the potential correlation between the structural features computed by SMAR ~~Scan®~~ SCAN and the S/MAR functional activity, the whole human genome has been analyzed with the combined method with very stringent parameters, in order to get sequences with the highest values for the theoretical structural features computed, which are called "super" S/MARs below. This was done with the hope to obtain predicted MAR elements with a very potential to increase transgene expression and recombinant protein production. The putative S/MARs hence harvested were first analyzed from the bioinformatics perpective in an attempt to characterize and classify them. --

On page 28, please amend the paragraph starting on line 3 as follows:

-- As whole human genome sequence, all human RefSeq (National Center for Biotechnology Information, The NCBI handbook [Internet]. Bethesda (MD): National Library of Medicine (US), Oct. Chapter 17, The Reference Sequence (RefSeq) Project, 2002 contigs (release 5) were used and analyzed with the combined method, using SMAR ~~Scan®~~ SCAN as filter in the first level processing, employing default settings except for the highest bend cutoff value, whereas a stringent threshold of 4.0 degrees (instead of 3.202 degrees) has been used for the DNA bending criterion. --

On page 29, please amend the paragraph starting on line 12 as follows:

-- The 1757 predicted "super" S/MARs sequences obtained previously by SMAR ~~Scan®~~ SCAN were then analyzed for potential transcription factors binding sites. This has been achieved using RMatch™ Professional (Kel AE, Gossling E, Reuter I, Cheremushkin E, KelMargoulis OV, Wingender E, MATCH: A tool for searching transcription factor binding sites in DNA sequences, *Nucleic Acids Res.* 31(13):35769, 2003), a weight matrixbased tool based on TRANSFAC (Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnhauser R, Pruss M, Schacherer F, Thiele S, Urbach S, The TRANSFAC system on gene expression regulation, *Nucleic Acids Research* , 29(1):2813, 2001). Match™ 2.0 Professional

has been used with most of the default settings Match™ analysis was based on TRANSFAC Professional, release 8.2 (20040630). The sums of all transcription factors binding prediction on the 1757 sequences analyzed according to Match™ are in Table 3. Based on this table, only the transcription factors totalizing at least 20 hits over the 1757 sequences analyzed were considered for further analyses. --

On page 31, please amend the paragraph starting on line 4 as follows:
-- The correlation between AT and TA dinucleotide percentages and the DNA highest bend as computed by SMAR ~~Scan®~~ SCAN is depicted in Fig. 17 for the predicted S/MAR sequences and in Fig.18 for the nonS/MAR sequences. The different scatterplots of these figures show that the TA percentage correlates well with the predicted DNA bend as predicted by SMAR ~~Scan®~~ SCAN. --

On page 32, please amend the paragraph starting on line 5 as follows:
-- In order to get an insight on S/MAR evolution, orthologous intergenic regions of human and mouse genomes have been analysed with SMAR ~~Scan®~~ SCAN. The data set used is composed of 87 pairs of complete orthologous intergenic regions from the human and mouse genomes (Shabalina SA, Ogurtsov AY, Kondrashov VA, Kondrashov AS, Selective constraint in intergenic regions of human and mouse genomes, *Trends Genet*, 17(7):3736, 2001) (average length ~12 000 bp) located on 12 human and on 12 mouse chromosomes, the synteny of these sequences was confirmed by pairwise sequence alignment and consideration of the annotations of the flanking genes (experimental or predicted). --

On page 32, please amend the paragraph starting on line 15 as follows:
-- Analysis of the 87 human and mouse orthologous intergenic sequences have been analysed with SMAR ~~Scan®~~ SCAN using its default settings. Analysis of the human sequences yielded a total of 12 S/MARs predicted (representing a total length of 4 750 bp), located on 5 different intergenic sequences. --

On page 32, please amend the paragraph starting on line 20 as follows:
-- Among the three human intergenic sequences predicted to contain a "super" S/MAR using SMAR ~~Scan®~~ SCAN stringent settings, one of the corresponding mouse orthologous intergenic sequence is also predicted to contain a S/MAR (human EMBL ID: Z96050, position 28 010 to 76

951 othologous to mouse EMBL ID: AC015932, positions 59 884 to 89 963). When a local alignement of these two orthologous intergenic sequences is performed, the best local alignement of these two big regions correspond to the regions predicted by SMAR ~~Scan®~~ SCAN to be S/MAR element. A manual search for the mouse orthologs of the two other human intergenic sequences predicted to contain a "super" S/MAR was performed using the Ensembl Genome Browser. The mouse orthologous intergenic sequences of these two human sequences were retrieved using Ensembl orthologue predictions (based on gene names), searching the orthologous mouse genes for the pairs of human genes flanking these intergenic regions. --

On page 32, please amend the paragraph starting on line 34 as follows:
-- Because SMAR ~~Scan®~~ SCAN has been tuned for human sequences and consequently yields little "super"MARs with mouse genomic sequences, its default cutoff values were slightly relaxed for the minimum size of contiguous hits to be considered as S/MAR (using 200 bp instead of 300 bp). Analysis by SMAR ~~Scan®~~ SCAN of these mouse sequences predicted several S/MARs having high values for the different computed structural features. This finding suggests that the human MAR elements are conserved across species. --

On page 34, please amend the paragraph starting on line 37 as follows:
-- To determine whether any aspects of DNA primary sequence might distinguish the active B, K and F regions from the surrounding MAR sequence, we analyzed the 5'-MAR with MAR Scan® SCAN. Of the 38 nucleosomal array prediction tools, three were found to correlate with the location of the active MAR sub-domains (Fig. 9A). Location of the MAR B, K and F regions coincides with maxima for DNA bending, major groove depth and minor groove width. A weaker correlation was also noted with minima of the DNA melting temperature, as determined by the GC content. Refined mapping over the MAR F fragment indicated that the melting temperature valley and DNA bending summit indeed correspond the FIB sub-fragment that contains the MAR minimal domain (Fig. 9B). Thus active MAR portions may correspond to regions predicted as curved DNA regions by this program, and we will refer to these regions as CUE-B, CUE-K and CUE-F in the text below. Nevertheless, whether these regions correspond to actual bent DNA and base-pair unwinding regions is unknown, as they do not correspond to bent DNA as predicted by MAR Wiz (Fig.9B). --

On page 40, please amend the paragraph starting on line 32 as follows:

-- **Example 16 : Use of MARs identified with SMAR ~~Scan®~~ SCAN II to increase the expression of a recombinant protein.** --

On page 40, please amend the paragraph starting on line 35 as follows:

-- Four MAR elements were randomly selected from the sequences obtained from the analysis of the complete human genome sequence with SMAR ~~Scan®~~ SCAN or the combined method. These are termed 1_6, 1_42, 1_68, (where the first number represents the chromosome from which the sequence originates, and the second number is specific to the predicted MAR along this chromosome) and X_S29, a "super" MAR identified on chromosome X. These predicted MARs were inserted into the pGEGFPControl vector upstream of the SV40 promoter and enhancer driving the expression of the green fluorescent protein and these plasmids were transfected into cultured CHO cells, as described previously (Zahn-Zabal, M., et al., *Development of stable cell lines for production or regulated expression using matrix attachment regions.* J Biotechnol, 2001. **87**(1): p. 29-42). Expression of the transgene was then analyzed in the total population of stably transfected cells using a fluorescent cell sorter (FACS) machine. As can be seen from Fig. 19, all of these newly identified MARs increased the expression of the transgene significantly above the expression driven by the chicken lysosyme MAR, the "super" MAR X_S29 being the most potent of all of the newly identified MARs. --